

# IntEnz, the integrated relational enzyme database

Astrid Fleischmann\*, Michael Darsow, Kirill Degtyarenko, Wolfgang Fleischmann, Sinéad Boyce<sup>1</sup>, Kristian B. Axelsen<sup>2</sup>, Amos Bairoch<sup>2</sup>, Dietmar Schomburg<sup>3</sup>, Keith F. Tipton<sup>1</sup> and Rolf Apweiler

EMBL Outstation—European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK, <sup>1</sup>The University of Dublin, Trinity College, Dublin 2, Ireland, <sup>2</sup>Swiss Institute of Bioinformatics (SIB), Geneva, Switzerland and <sup>3</sup>Cologne University, Cologne, Germany

Received August 15, 2003; Revised and Accepted October 20, 2003

## ABSTRACT

**IntEnz is the name for the Integrated relational Enzyme database and is the official version of the Enzyme Nomenclature. The Enzyme Nomenclature comprises recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) on the nomenclature and classification of enzyme-catalysed reactions. IntEnz is supported by NC-IUBMB and contains enzyme data curated and approved by this committee. The database IntEnz is available at <http://www.ebi.ac.uk/intenz>.**

## INTRODUCTION

The explosion of structural data that has resulted from gene-sequencing and proteomic studies has emphasized the need to relate structures to functions. It is important that these structural data are integrated with the functional data on catalytic proteins that are available from a number of sources, such as the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) Enzyme List (1), ENZYME (2) and BRENDA (3) databases. The Enzyme List represents a well established and officially recognized functional classification system.

The Integrated relational Enzyme database IntEnz provides a complete, freely available database storing the most up-to-date version of the Enzyme List, which contains enzyme data curated and approved by the NC-IUBMB.

In addition tools have been developed to maintain the database and allow propagation of new and updated biochemical terminology, which has been integrated into the database. These tools are connected to the ChEBI database, which provides a definitive dictionary of compounds, to improve the quality of the IntEnz vocabulary. ChEBI stands for dictionary of Chemical Compounds of Biological Interest. The ChEBI database is also hosted at the European Bioinformatics Institute (EBI), but is still under development.

Enzyme data will be represented in a number of different ways, one of which is the official NC-IUBMB view, which has been developed and can be accessed at <http://www.ebi.ac.uk/intenz>.

## CLASSIFICATION OF ENZYMES

The NC-IUBMB classification and nomenclature of enzymes, commonly known as the Enzyme List, differs from most biochemical classification systems in that it is based on function, i.e. the reaction catalysed, rather than structure. The basis of the system is the assignment of a specific numerical identifier, the EC number, which identifies the enzyme in terms of the reaction catalysed. The first digit represents the type of reaction catalysed. At present six such classes of reaction are recognized, as summarized in Table 1. The second digit of the EC number refers to the subclass, which generally contains information about the type of compound or group involved. The third digit, the sub-subclass, further specifies the nature of the reaction and the fourth digit is a serial number that is used to identify the individual enzyme within a sub-subclass. The classification of enzymes is described in detail elsewhere (4,5).

Although several distinct proteins may catalyse the same reaction, they would all be ascribed the same EC number, since the Enzyme List is a system based upon the reaction catalysed. Information about structural diversity is obtained by linkage of enzymes in the List to related entries in other databases. The information for each enzyme within the Enzyme List is contained in seven distinct fields, as summarized in Table 2. There is also a glossary of compound names, which may be used to search the Enzyme List. The Enzyme List is maintained by Trinity College Dublin on behalf of the NC-IUBMB. Suggestions for new entries or for modifications to existing entries can be found on the IntEnz homepage.

## DATA IN IntEnz

IntEnz is supported by the NC-IUBMB and contains enzyme data that is curated and approved by the Nomenclature Committee. The relational database (IntEnz) implemented and supported by the EBI is the master copy of the Enzyme Nomenclature data. The goal is to create a single relational database containing enzyme data from three different sources.

(i) Trinity College Dublin (TCD) maintains on behalf of the NC-IUBMB the Enzyme List and is involved in many other aspects of biochemical nomenclature.

\*To whom correspondence should be addressed. Tel: +44 1223 494444; Fax: +44 1223 494468; Email: Astrid.Fleischmann@ebi.ac.uk

(ii) The Swiss Institute of Bioinformatics (SIB) produces an Enzyme Nomenclature database (ENZYME).

(iii) The University of Cologne produces an enzyme function database (BRENDA), which contains a large amount of information on substrates, products and inhibitors.

The EBI has already implemented a relational database version storing the most up-to-date version of the Enzyme List, which contains enzyme data curated and approved by NC-IUBMB. Figure 1 shows a screenshot of the classification of EC 1.1.1.51.

## DATABASE DEVELOPMENT

The database is implemented using a relational database management system (ORACLE) and consists of development, production and public versions of the database.

Web servers build a bridge between the database server machines and the users on the internet. They accept user requests for data, interrogate the appropriate database, summarize the result in a form fit for human consumption and deliver it to the user's internet browser.

**Table 1.** The six reaction classes used in the enzyme list

Class	Name	Reaction catalysed
1	Oxidoreductase	$AH_2 + B = A + BH_2$ or $AH_2 + B^+ = A + BH + H^+$
2	Transferase	$AX + B = A + BX$
3	Hydrolase	$AB + H_2O = AH + BOH$
4	Lyase	$A=B + X-Y = A-B$      X  Y
5	Isomerase	$A = B$
6	Ligase	$A + B + NTP^a = A-B + NDP + P$ or $A + B + NTP = A-B + NMP + PP$

Adapted from (4).

<sup>a</sup>NTP, nucleoside triphosphate.

The web application consists of a back- and a front-end. The back-end is responsible for database-related functionality (retrieving and updating) and processing of data. It is connected with the production and public database to allow access either for adding/modifying or for retrieving data. Representing data and providing an interface for user interaction are the fundamental features of the front-end. The interface is compatible with all common browsers (e.g. Microsoft Internet Explorer, Netscape Mozilla, OmniWeb, Safari). The front-end is implemented using HTML and JavaServerPages. The functionality of the back-end is mainly achieved using Java Servlets.

## PUBLIC AND PRODUCTION DATABASE

The users of the web application are categorized into two groups: the general user and the IntEnz curator. The general user has access to the public database to retrieve information of interest, but will not be able to modify or add data. The IntEnz curator is able to access the production database to perform changes to the data.

The reasons for the separation of production data and publicly accessible data are as follows.

(i) Security: only IntEnz curators access the data on the production database. The public access the public web server and the public database, which hold a copy of the production data.

(ii) Fail over: in emergencies and during scheduled maintenance sessions, one set of machines can serve both the production and public versions of the database, albeit at a reduced performance rate.

(iii) Technical: there are different usage patterns for production and public machines. Usage of the production machines can be split into interactive use during European office hours and bulk updates during the night, while the public machines have lower, but more constant access rates to serve the world wide web.

**Table 2.** Fields used in the IUBMB Enzyme List<sup>a</sup>

Field	Description
EC number	The classification number, which is made up of four groups of digits, identifies the enzyme by the reaction catalysed.
Common name	The most commonly used name for the enzyme is usually used, provided that it not ambiguous or misleading. A number of generic words indicating reaction types are commonly used, such as dehydrogenase, reductase, oxidase, peroxidase, kinase, tautomerase, deaminase, dehydratase, etc.
Reaction	The actual reaction catalysed is written, where possible, in a general form of the type: $A + B = P + Q$ . It gives no indication of the equilibrium position of the reaction or, indeed, whether it is readily reversible. The direction chosen for the reaction and systematic name is the same for all the enzymes in a given class, even if this may not be the preferred direction for all of them. Where an enzyme has a wide specificity, compound names may be replaced by class names, for example the reaction catalysed by alcohol dehydrogenase (EC 1.1.1.1) is written as: an alcohol + $NAD^+$ = an aldehyde or ketone + $NADH + H^+$ .
Other name(s)	Any other names that have been used for the enzyme. This is to be as comprehensive a list as possible to aid searching for any specific enzyme. The inclusion of a name in this list does not mean that its use is encouraged. In cases where the same name has been given to more than one enzyme, this ambiguity will be indicated.
Systematic name	This describes the reaction catalysed in unambiguous terms. It consists of two parts. The first contains the name of the substrate or substrates (each separated by a colon). The second part, ending in -ase, indicates the nature of the reaction. Generic words are used to indicate the type of reaction, e.g. oxidoreductase, oxygenase, transferase (with a prefix indicating the nature of the group transferred), hydrolase, lyase, racemase, epimerase, isomerase, mutase, ligase. If additional information is needed to make the reaction clear, a phrase indicating the reaction or a product is added in parentheses after the second part of the name, e.g. (ADP-forming), (dimerizing), (CoA-acylating).
Comments	Brief comments on the nature of the reaction catalysed, possible relationships to other enzymes, species differences, metal-ion requirement, etc.
References	Key references on the identification, nature, properties and function of the enzyme.

<sup>a</sup>Note: the procedures used for the nomenclature and classification of the peptidases (EC 3.4.-.-) differ from those used for other enzymes in some respects.

The screenshot shows a web browser window displaying the EMBL-EBI Integrated Enzyme Database (IntEnz) entry for EC 1.1.1.51. The browser's address bar shows the URL: <http://www.ebi.ac.uk/intenz/query?cmd=SearchID&id=1050>. The EMBL-EBI logo is at the top, with navigation links for Home, About EBI, Research, Services, Toolbox, Databases, Downloads, and Submissions. The IntEnz logo is on the left side of the page.

**EC 1      Oxidoreductases**  
**EC 1.1    Acting on the CH-OH Group of Donors**  
**EC 1.1.1   With NAD<sup>+</sup> or NADP<sup>+</sup> as acceptor**  
**EC 1.1.1.51   3(or 17)β-hydroxysteroid dehydrogenase**

**IUBMB Enzyme Nomenclature**  
**EC 1.1.1.51**

**Common name:** 3(or 17)β-hydroxysteroid dehydrogenase

**Reaction:** testosterone + NAD(P)<sup>+</sup> = androst-4-ene-3,17-dione + NAD(P)H + H<sup>+</sup>

**Systematic name:** 3(or 17)β-hydroxysteroid:NAD(P)<sup>+</sup> oxidoreductase

**Other name(s):** 17-ketoreductase  
 17β-hydroxy steroid dehydrogenase  
 3β-hydroxy steroid dehydrogenase  
 3β-hydroxysteroid dehydrogenase  
 β-hydroxy steroid dehydrogenase

**Comments:** Also acts on other 3β- or 17β-hydroxysteroids. cf. EC 1.1.1.209 3(or 17)α-hydroxysteroid dehydrogenase.

**Links to other databases:** [BRENDA](#), [WIT](#), [EXPASY](#), [KEGG](#), [GTD](#)

**CAS registry number:** 9015-81-0

**References:**

1. Dahm, K. and Breuer, H. Anreicherung einer 17β-hydroxysteroid:NAD(P)-oxydoreduktase aus der Nebenniere der Ratte. *Hoppe-Seyler's Z. Physiol. Chem.*, 336, (1964) 63-68.
2. Lynn, W.S. and Brown, R.H. The conversion of progesterone to androgens by testes. *J. Biol. Chem.*, 232, (1958) 1015-1030.
3. Marcus, P.I. and Talalay, P. Induction and purification of α- and β-hydroxysteroid dehydrogenases. *J. Biol. Chem.*, 218, (1956) 661-674.
4. Schultz, R.M., Groman, F.V. and Engel, L.L. 3(17)β-Hydroxysteroid dehydrogenase of *Pseudomonas testosteroni*. A convenient purification and demonstration of multiple molecular forms. *J. Biol. Chem.*, 252, (1977) 3775-3783. [Medline UI: 77187825]
5. Talalay, P. and Dobson, M.M. Purification and properties of a α-hydroxysteroid dehydrogenase. *J. Biol. Chem.*, 205, (1953) 823-837.

**[EC 1.1.1.51 created 1961]**

Please contact [support@ebi.ac.uk](mailto:support@ebi.ac.uk) with any problems or suggestions regarding this site.

**Figure 1.** Screenshot of the Enzyme List entry for EC 1.1.1.51.

(iv) Scientific: new and updated data on the production machines needs to be kept separate until it is peer reviewed by the NC-IUBMB who are in editorial control of the data. Only then can it be copied to the public database and public web servers. This process is crucial to maintaining the consistency of the database through editorial control by the NC-IUBMB.

The building of such a robust database and tools environment automates database updating as much as possible. It also provides the general user and the IntEnz curator with stable search and curation tools.

## SEARCH TOOLS

The database is provided with an advanced search tool that enables users to exploit the data contained in the database. The search tool permits the user to:

- (i) limit searches to defined fields;
- (ii) combine search terms or search results with Boolean operators;
- (iii) use word expansion operators, e.g. the wildcard operator or the fuzzy operator.

## FUTURE DEVELOPMENT

IntEnz will be further populated by corresponding data from the ENZYME database of the SIB and from the BRENDA database of the University of Cologne. The discrepancies between the equivalent data fields from these data sets are to be resolved as the next step towards a controlled vocabulary for biochemical terminology. These common data comprise the EC number, common name of an enzyme, systematic

name of an enzyme, other name(s) of an enzyme, reaction catalysed and comments.

The representation of enzyme data will be done in a number of different ways. We already have a NC-IUBMB authorized view of the data. After resolving the discrepancies between the equivalent data fields we will also be able to retrieve a flat-file output of IntEnz, which will replace the ENZYME.

## ACKNOWLEDGEMENTS

We thank Nicola Mulder, Maria Krestyaninova, Sandra Orchard and Bob Vaughan who helped us with the annotation of entries. We are also grateful for the support from David Binns, Christian Gran, Henning Hermjakob, Alexander Kanapin, Rodrigo Lopez and Muruli Rao. The IntEnz project is supported by the BioBabel grant (no. QLRT-2000-00981) of the European Commission.

## REFERENCES

1. IUBMB (1992) *Enzyme Nomenclature: Recommendations (1992) of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology*. Academic Press, San Diego, CA.
2. Bairoch, A. (2000) The ENZYME database in 2000. *Nucleic Acids Res.*, **28**, 304–305.
3. Schomburg, I., Chang, A. and Schomburg, D. (2002) BRENDA, enzyme data and metabolic information. *Nucleic Acids Res.*, **30**, 47–49.
4. Boyce, S. and Tipton, K.F. (2000) History of the enzyme nomenclature system. *Bioinformatics*, **16**, 34–40.
5. Tipton, K.F. and Boyce, S. (2000) Enzyme classification and nomenclature. In *Nature Encyclopedia of Life Sciences*. Nature Publishing Group, London. <http://www.els.net/> [doi:10.1038/npge.els.0000710]